6.4

THE USE OF AVIATION SYSTEM PERFORMANCE METRICS IN QUALITATIVE FORECAST EVALUATION

Matthew Lorentson *
National Weather Service, Silver Spring, Maryland

1. INTRODUCTION

The National Weather Service (NWS) uses 'probability of detection' (POD) and 'false alarm ratio' (FAR) to assess forecast accuracy of many products and services, including aviation forecasts like TAFs—Terminal Aerodrome Forecasts (National Weather Service 2011). POD and FAR provide helpful quantitative verification information about forecast categories, such as cloud cover and ceiling height, visibility, wind direction and speed, etc. Each basic category can have a separate and significant impact on aviation operations, so it is useful to differentiate between them.

Evaluation of meteorological forecast phenomena becomes complex when categories and impacts are not isolated, such as in the case of "convection", or thunderstorms. The presence of deep convection can be verified with equipment like radar and lightning detection sensors, but it is not a category that specifically describes singular aviation impacts. Thunderstorms possess several threat characteristics, any of which can alternately cause the greatest impact on aviation operations at a moment in time. For example, a small, isolated, relatively low-topped thunderstorm within ten miles of an airport may impact en route aviation operations to some degree, but also generate a significant surface wind speed and direction change at the nearby terminal. Such an occurrence could have a major impact on aviation operations, or almost no impact, depending on the airport's runway configuration, volume of traffic, flight mix, wind detection equipment, air traffic planning, etc. To reduce the complexity associated with multiple variables, it makes sense to evaluate weather forecasts in terms of basic categories and avoid generalized criteria like convection.

According to Steiner et al. (2009), weather forecast information needs to be fully integrated into the Traffic Flow Management (TFM) decision-making process—translated into aviation impact forecasts—in order to be most effective. This is one of the key goals for both the Next Generation Air Transportation System and the Single European Sky ATM Research. As forecasts already are an element of TFM success, it seems reasonable now, and in the future, to evaluate forecasts in terms of TFM impact.

* *Corresponding author address:* Matthew Lorentson, NWS OCWWS, Aviation Services Branch, SSMC2, 1325 East West Hwy., Silver Spring, MD 20910-3283; e-mail: matthew.lorentson@noaa.gov

During the past few years, "delays" have been used by the NWS as an informal measure of forecast impact on aviation operations. While this seems intuitively reasonable, delays do not measure forecast quality directly and are a broad aviation impact category. Evans et al. (2004) proposed that direct comparison of delays before and after a forecast system is introduced is very hard to carry out in practice, even though it appears quite straightforward, and that a "decision/modeling" [qualitative] method is the only feasible way to assess potential improvement in system products.

This is undoubtedly true in the analysis of generalized forecast categories, e.g., convection. However, we know that forecasts influence TFM decisions and impacts, so it should theoretically be possible to reduce forecast information parameters and impact variables to a sufficient degree that a significant correlation exists between the two. This concept can be applied to ongoing daily and seasonal evaluation of forecast quality and impact. Pfleiderer et al. (2007) assert that, when making forecast system deployment decisions prospectively, projections should be validated against objective metrics following implementation, specifically with regard to metrics that describe efficiency. Therefore, if measures like delays are used in forecast evaluation, an efficiency metric should be included in the analysis to at least account for history effects.

Aviation System Performance Metrics (ASPM) provide information about TFM and Air Traffic Control (ATC) performance that can be used to supplement/enhance delay information. ASPM Efficiency Rates were specifically created to measure an ATC facility's ability—and by extension, that of the air traffic system—to do what it says in can do (Pfleiderer et al. 2007). Together with broader measures of TFM performance, such as ASPM "Percent on Time Arrivals" or Air Traffic Operations Network (OPSNET) "Weather Delays," this kind of efficiency metric should aid in the evaluation of forecast performance. The goal of this study is to demonstrate, via 2010 San Francisco stratus season forecasts, the potential use of ASPM and OPSNET Weather Delays as objective metrics and predictors of specific, isolated forecast information.

2. DELAYS AND ASPM

The ASPM database is used by the Federal Aviation Administration (FAA) to document and evaluate TFM performance. The metrics are demand-based. Demand is defined as the number of aircraft that intend to land at a specific airport, or are ready to depart from the airport, by unit of time (Federal Aviation Administration 2011a).

ASPM combines data from the FAA en route system (Host) on aircraft positions, flight plans, Official Airline Guide schedules, Airline Service Quality Performance, and (for some of the major carriers) Out/Off/On/In (OOOI) data. OOOI data consists of:

- Actual gate departure time (“Out”)
- Actual flight takeoff time (“Off”)
- Actual flight landing time (“On”)
- Actual gate arrival time (“In”)

Hence, it is possible to make much more detailed, objective, quantitative studies of where a flight delay occurred than is the case with OPSNET delays (Evans et al. 2004). Three ASPM types—two efficiency ratings and an arrivals metric—were chosen to supplement OPSNET Weather Delay information in this study.

2.1 *ASPM Percent on Time Arrivals*

The Percent on Time Arrivals (POTA) metric used in this study simply describes the number of flights that arrive at a gate less than 15 minutes past the flight plan time in relation to the total number of Arrivals for Metric Computation—in other words, the number of on time arrivals divided by total arrivals. The last scheduled flight plan before Wheels Off time is used for metric calculations (Federal Aviation Administration 2011b).

2.2 *ASPM Demand-based Efficiency Rates*

[Efficiency Rate definitions/examples derived from Wine (2005) and Diana (2005).]

Arrival Efficiency Rate is used to determine how well demand for arrivals is met. It is the number of actual airport arrivals divided by the lesser of arrival demand or the airport arrival rate. These terms are defined as:

- Actual arrivals: how many aircraft landed during a quarter hour
- Arrival demand: how many aircraft wanted to land during that quarter hour
- Airport Arrival Rate (AAR): the facility-set number of aircraft expected to land during that quarter hour

Demand is not derived from scheduled traffic, but in the following manner for each flight:

- Start of demand = wheels-off time + filed en route time
- End of demand = wheels-on time (touch-down)

An example of the Arrival Efficiency Rate calculation is provided in Fig. 1.

Departure Efficiency Rate is the number of actual departures divided by the lesser of departure demand or the airport departure rate. These terms are defined as:

- Actual departures: how many aircraft departed during a specific quarter hour
- Departure demand: how many aircraft wanted to depart during that quarter hour
- Airport Departure Rate (ADR): the facility-set number of aircraft expected to depart during that quarter hour

Departure demand is not derived from the scheduled traffic, but in the following manner for each flight:

- Start of Demand = gate-out time + unimpeded taxi-out time
- End of Demand = wheels-off time (lift-off)

Fig. 2 shows an example of the Departure Efficiency Rate calculation.

System Airport Efficiency Rate (SAER) is a weighted average of the Arrival Efficiency Rate and Departure Efficiency Rate. Arrivals and departures for SAER are calculated in terms of aircraft demand per quarter-hour increment. An example of the SAER calculation is provided (Fig. 3).

Terminal Airport Efficiency Rate (TAER) is similar to SAER, but arrival demand starts at 100 miles from the airport and compares an average wheels-on time vs. actual wheels-on time. TAER is reported for a time period of an hour or more. For reports covering a period greater than an hour, the reported TAER is a weighted average of every hour in the period (Federal Aviation Administration 2011a).

Theoretically, Efficiency Ratings should provide relative information about TFM performance that can help isolate forecast quality. For example, if the POTA measure is fairly average for a specific weather scenario, terminal, and day, but the SAER and TAER were very low for the given situation, one might expect to find that high quality forecast information helped limit the number of delayed arrivals. The hypothesis that SAER or TAER enhance broader TFM metrics like Weather Delays and POTA in the prediction of forecast quality will be tested with statistical regression in Section 4.

2.3 *OPSNET Weather Delays*

The FAA’s OPSNET database contains delay causality information, unlike that described by POTA and other ASPM information. Delays are assigned to five major categories within OPSNET: Weather, Volume, Equipment, Runway, and Other. This is a major strength, but there are some major deficiencies associated with OPSNET delays too; delays are reported by humans at ATC facilities and can be inaccurate, are excluded if initiated by pilot/airlines, and are not reported if less than 15 minutes. Delay reporting methods are also subjective and differ by facility. (Evans et al. 2004).

Section 8.f.1 of FAA Order 7210.55F (2011d) describes a list of impact conditions, or constraints, used to define/assign delays to the Weather category. This list of impact conditions includes low ceilings, the meteorological condition of concern used in this research.

3. DATA SAMPLE

Unfortunately, there are few situations where a single meteorological condition creates consistent impacts on aviation. However, certain operationally significant weather phenomena do occur on a regular basis and at relatively regular times, such as stratus ceilings at San Francisco (SFO). Low stratus at SFO negatively affects the ability of ATC to operate and TFM to plan effectively. Stratus ceilings are disruptive to flights on approach to the airport; the reduction in visibility at relatively low altitudes requires greater spacing between flights for safety reasons. This need for space reduces the number of flights that can land at the airport per hour and usually causes delays. There is immediate improvement in arrival capacity and capability when the stratus ceiling lifts or clears out of the approach corridor.

The 2010 stratus season exhibited some 72 typical “Stratus Days” for which clearing times were predicted, observed, and recorded by the Oakland Air Route Traffic Control Center (ZOA ARTCC) Center Weather Service Unit (CWSU). The raised acceptance rate, or moment when the AAR was increased by FAA Traffic Management to accommodate inbound flight demand, was used to define the actual stratus improvement time. AAR is a dynamic parameter specifying the number of arrival aircraft that an airport, in conjunction with terminal airspace, can accept under specific conditions throughout any consecutive sixty-minute period (Federal Aviation Administration 2011c). AAR’s are raised at SFO when pilots can maintain visual separation between aircraft and when Air Traffic Control can accommodate a greater number of arrivals. Raised AAR represents the stratus improvement condition that the CWSU forecasts.

An initial clearing forecast is provided to TFM decision makers by 12:45 UTC each Stratus Day, when it is used in a planning teleconference. The CWSU forecaster differentiates between “typical” and “non-typical” stratus events; a typical stratus event occurs when on-shore (west) winds prevail, without the presence of precipitation or synoptically driven influences like fronts. Low ceilings reduce airport capacity significantly, but operations remain in the optimal west-approach runway configuration. Essentially, just one weather condition/category impacts the AAR on typical Stratus Days.

The data sample used for evaluation here (Fig. 4) includes the 12:45 UTC “CWSU Forecast“ clearing time and corresponding ASPM/OPSNET metrics for each SFO Stratus Day in 2010. The CWSU Forecast data represents a forecast error time in minutes; this time is the difference between forecast clearing time and actual clearing time (raised AAR). Positive CWSU Forecast numbers describe an actual clearing time that occurred earlier than forecast; negative numbers mean the stratus deck lingered over the Bay Area longer than expected by the forecaster. It is assumed that a smaller error or smaller difference between forecast clearing time and actual clearing time equates to a more accurate, higher quality forecast. This is particularly true with regard to negative CWSU Forecast numbers. However, large, positive CWSU Forecast errors should not typically result in more weather delays; when the stratus deck clears relatively early, ATC can accommodate more flights and incur fewer delays. En route flights that arrive from the East Coast during the late morning hours can land without the need for an accurate forecast. Efficiency rating metrics that describe TFM performance are again an important element of the analysis here; large delay numbers on large *positive* forecast error days would presumably be an indicator of low efficiency. While a clearly communicated and accurate forecast should enhance the chances for efficient operations, there are a multitude of other variables not related to forecast interpretation that can cause Weather Delay numbers to increase. Among these are traffic complexity, facility staffing, and ability of planners to reach consensus, all of which can potentially influence efficiency.

4. DATA ANALYSIS

4.1 *Descriptive Statistics*

A boxplot of Weather Delays (Fig. 5) shows the positive skewness of the sample, with a long whisker from around 90 to 150 and short whisker from 0 to 30. The box represents the inter-quartile range, with a median value of just over 50 Weather Delays and a mean value (red dashed line) of just over 60, which also emphasizes the positive skew. There is relatively low variability; the majority of days are within the 30 to 90 Weather Delay range. The sample is heavy-tailed, with three moderate outliers (two of them overlap in Fig. 5) and one extreme outlier. These outliers (Fig. 6) are unusual, but appear to be an important aspect of the distribution. The CWSU Forecast was over 300 minutes in error on the two largest weather delay days. SAER is around 90 on the other two outlier days, a relatively low efficiency rating with regard to this data sample that appears to explain the two remaining outlier values. These four outliers belong with the other data values; their removal from the data sample regression reduces correlation significantly.

4.2 *Linear Regression*

[This approach is similar to that used by Berk and Carey 2000, 299—367.]

A scatterplot of CWSU Forecast error and Weather Delays (Fig. 7) highlights data relationships. Generally, forecast clearing times are within 100 minutes of actual clearing times and Weather Delays are below 150 on typical Stratus Days. Early clearing times, +100 minutes relative to forecast clearing times, tend to correlate with lower Weather Delay numbers. ATC is able to recover and deliver more arrivals to the airport during peak morning traffic times on these days.

Note that on days when the greatest number of Weather Delays occurred, stratus cleared very late in relation to the forecast expectation. The scatterplot of Weather Delays in Fig. 7 helps illustrate the linear relationship and negative correlation with CWSU Forecast error. The trend line for this data set crosses the x-axis at approximately 73; thus, if a linear relationship were representative for the 2010 season, a perfect Stratus Day clearing forecast would result in about 73 Weather Delays, given “average” ATC efficiency.

One of the goals of statistical regression is prediction. To evaluate the use of weather delays as a predictor of forecast quality, a regression can be used to calculate the least-squares estimates (Fig. 8). The regression of CWSU Forecast and Weather Delay data samples yields an R^2 value (coefficient of determination) of 0.344, which indicates that 34.4% of the variation in CWSU Forecast quality can be explained by the change in Weather Delays. From the Analysis of Variance (ANOVA) table, one (1) degree of freedom is attributed to the regression and that the sum of squares attributed to residuals is nearly double that of the regression. An F-ratio of 36.76 and P-Value of 6×10^{-8} indicate that the regression is statistically significant.

A plot of the residuals (differences between the observed CWSU Forecast values and the regression-predicted numbers) can be used to verify that this data does not violate normal distribution. The plot of residuals (Fig. 9) indicates that, although the points do not fall perfectly on the line, the departure is not strong enough to invalidate our assumption of normality.

Fig. 10 shows a correlation matrix of the full data sample. CWSU Forecast correlates meaningfully with On Time Arrivals and Weather Delays, with Multiple R statistics of 0.617 and -0.587 respectively. Each has a Pearson Probability (P-Value) less than 0.05. Meanwhile, the Airport Efficiency Ratings do not show a meaningful correlation with CWSU Forecast, which is a desired result. A scatterplot matrix (Fig. 11) illustrates the correlation matrix relationships. CWSU Forecast data exhibits a roughly linear relationship with both POTA and Weather Delays. The strongest linear relationship lies between POTA and Weather Delays; this correlation stands out clearly within the matrix and has a Multiple R of -0.875 (R-square 0.766).

A multiple regression of all four metric categories together with CWSU Forecast yields improvements in the statistical relationship (Fig. 12), an R-square value of 0.46—46% of the variance in CWSU Forecast error is due to the differences in these predictors. The prediction equation, derived from the Coefficients column, is:

CWSU Forecast = 793.08 + 2.52(TAER) – 14.01(SAER) + 4.81(POTA) – 0.53(Weather Delays)

SAER, while not a meaningful predictor of CWSU Forecast quality as a single variable, becomes the most important predictor in the multiple regression calculation, with a P-value of .0006. The t-stat, or t-value, for SAER is -3.59. The probability of a t-value this large or larger in absolute value is about .0006, well below the 0.05 significance level. So we reject the null hypothesis that the coefficient is 0 at the 5% level and accept the alternative hypothesis. Note that only the coefficients for SAER and POTA have significant P-values. Therefore, we should not focus this analysis on TAER and Weather Delays.

A regression of SAER and POTA (Fig. 13) produces similar results to the four-coefficient multiple regression, with an adjusted R-square of 0.45 and stout significance (P-value 3×10^{-10}). Now, a test for regression assumptions:

1. A plot of the CWSU Forecast values against the predicted values (Fig. 14) shows the success of the regression. The regression fits the data.

2. A plot of predicted values and residuals of the CWSU Forecast values (Fig. 15) magnifies the vertical spread of the data and verifies the regression assumptions. No curved pattern is apparent, and there is no trend that might suggest non-constant variance.

3. A plot of residuals against the individual predictor variables (Fig. 16 and 17) can reveal problems that are not clear from the regression predicted residuals in Fig. 15. It appears that there are no curves or trends with regard to POTA, although the SAER data appears to narrow somewhat as values decrease. This impression might be lessened if the outlier value ~84 is removed. To test non-constant variance, we can perform an arcsine-square root transformation on SAER:

Transformed SAER = $\sin^{-1}$ (SQ RT of SAER/100)

The transformed SAER variables yield similar results in the regression (Fig. 18) and perhaps help make the variance a little more constant. However, each point in the SAER score is worth -592.731 points in the arcsine of the square root of the CWSU Forecast error divided by 100; not a favorable simplification. The transformed regression is useful to validate the original regression. Since the variance is relatively constant in the transformation and it gives essentially the same results as the original regression, then the original results are valid.

4. A normal plot (Fig. 19) reveals residuals that do not fall perfectly on the line, similar to the normal plot of the full multiple regression with all four coefficients (Fig. 20). Although there is some departure, it does not appear strong enough to invalidate the assumption of normality.

With the final multiple regression, non-significant variables have been eliminated and the number of coefficients is reduced to two with very little reduction in statistical significance. This simple, linear equation can be used to help verify the quality of CWSU stratus forecasts from 2010, given the POTA and SAER at SFO:

CWSU Forecast = 721.81 – 12.537(SAER) + 6.904(POTA)

5. SUMMARY

It is difficult to quantify the impact that generalized weather phenomena have on operations within the National Airspace System. Thus, weather conditions should be reduced to simple impact categories in order to isolate significant correlations between forecast quality and TFM performance metrics. Weather Delays, the only TFM impact measure currently referenced by the NWS, may not be the best tool available to inform qualitative forecast evaluation. It is reasonable to expect that ATC efficiency metrics will enhance the information provided by broader measures of aviation system performance like Weather Delays or POTA. In the example of 2010 SFO CWSU stratus ceiling clearing forecasts, Weather Delays predicted approximately 35% of the CWSU Forecast quality variance, while the differences in SAER and POTA improved that predictability of variance to just over 45%.

6. REFERENCES

Berk, K. and P. Carey, 2000: *Data Analysis with Microsoft Excel*. Duxbury Press, 372 pp.

Diana, T., 2005: Presentation of Key Airport System Performance Metrics. [Available at ftp://apo130/Airport_Efficiency/ASPM_Key_Efficiency Metrics(3).ppt

Evans, J., S. Allan, and M. Robinson, 2004: Quantifying Delay Reduction Benefits for Aviation Convective Weather Decision Support Systems. *Proc. 11th Conference on Aviation, Range, and Aerospace Meteorology,* Hyannis, MA, Amer. Meteor. Soc. [Available online at http://www.ll.mit.edu/mission/aviation/publications/publication-files/WW-11278_ConvWxBenefits_FINAL.pdf]

Federal Aviation Administration, cited 2011a: APM Efficiency: Definitions of Variables. [Available online at http://aspmhelp.faa.gov/index.php/APM_Efficiency:_Definitions_of_Variables]

Federal Aviation Administration, cited 2011b: APM Weather Factors: Definitions of Variables. [Available online at http://aspmhelp.faa.gov/index.php/APM_Weather_Factors:_Definitions_of_Variables]

Federal Aviation Administration, cited 2011c: Order 7210.3C. [Available online at http://www.faa.gov/air_traffic/publications/atpubs/fac/1007.html]

Federal Aviation Administration, cited 2011d: Order 7210.55F. [Available online at http://www.faa.gov/documentLibrary/media/Order/7210.55FBasic.pdf]

National Weather Service, cited 2011: Operations and Services Performance, NWSPD 10-16 Verification Procedures. [Available online at http://www.weather.gov/directives/sym/pd01016001curr.pdf]

Pfleiderer, E., S.Goldman, and T. Chidester, 2007: Time Series Analyses of Integrated Terminal Weather System Effects on System Airport Efficiency Ratings. DOT Tech. Rep. DOT/FAA/AM-07/28 accomplished under approved task AM-B07-HRR-521, 28 pp. [Available online at http://www.dtic.mil/cgi-bin/GetTRDoc?Location=U2&doc=GetTRDoc.pdf&AD=ADA475572]

Steiner, M., R. Bateman, J. Krozel, Y. Liu, D. Megenhardt, M. Pocernich, and M. Xu, 2009: Translation of Ensemble Weather Forecasts into Probabilistic Air Traffic Capacity Impact. *Air Traf. Cont. Quar.,* **18(3),** 229—254.

Wine, C., 2005: Presentation of System Airport Efficiency Rate (SAER) and Terminal Arrival Efficiency Rate (TAER). [Available at http://aspm.faa.gov/aspm/ASPMframe.asp]

6. ACKNOWLEDGEMENTS

The author gratefully acknowledges NWS Aviation Weather Center Support Branch Chief David Bright and NWS Performance Branch Chief Doug Young for their edits and NWS ZOA CWSU Meteorologist-In-Charge Ken Venzke for providing the CWSU Forecast data sample.

7. ILLUSTRATIONS AND TABLES

Arrival Demand
Wheels-off: 9:55
Estimated time en route: 2:00 hours (120 minutes)
Wheels-on: 12:05

Start of demand: 9:55 + 120 minutes = 11:55
End of demand: 12:05

Demand occurs in two separate quarter-hour periods (11:45-11:59 and 12:00-12:14). Only one quarter-hour period is assigned for demand durations shorter than 15 minutes. If demand duration is split between quarters, the start of demand is shifted forward (in this case, to the 12:00 - 12:14 period.

Arrival Efficiency Rate
Actual arrivals/arrival demand (not to exceed AAR)
Example: BOS, December 6, 2003, from 11:45 to 11:59
Actual arrivals = 6
Arrival demand = 7
AAR = 9
Arrival Efficiency Rate = 6/7 = 85.71

Figure 1: Arrival Efficiency Rate Calculation

Departure Demand
Gate-out: 11:05
Unimpeded Taxi-Out Time: 12 minutes
Wheels-off: 11:55

Start of demand: 11:05 + 12 minutes = 11:17
End of demand: 11:55

Demand for this flight occurs in three quarter-hour periods (11:15 to 11:29, 11:30 to 11:44, and 11:45 to 11:59).

Departure Efficiency Rate
Actual departures/departure demand (not to exceed ADR)
Example: ORD, December 6, 2003, from 13:30 to 13:44
Actual departures = 21
Departure demand = 25
ADR = 27
Departure Efficiency Rate = 21/25 = 84.00

Figure 2: Departure Efficiency Rate Calculation

Arrival Adjustment
Wheels-off: 10:35
Estimated time en route: 1:50 (110 minutes)
Ground delay: 30 minutes
Wheels-on: 12:25

Start of Demand = 10:35 + 110 minutes – 30 minutes = 11:55
End of Demand = 12:25

Arrival demand occurs in three separate time periods (11:45-11:59, 12:00-12:14, and 12:15-12:29). Ground delay adjustment has a negative impact on arrival score.

Departure Adjustment
Gate-Out: 11:05
Unimpeded Taxi-Out Time: 12 minutes
Ground delay: 30 minutes
Wheels-off: 11:47

Start of Demand: 11:05 + 12 minutes + 30 minutes = 11:47
End of Demand: 11:47
Departure demand for this flight occurs in 1 time period: 11:45 to 11:59, but ground delay adjustment removes any negative impact on departure score.

System Airport Efficiency Rate - SAER
Arrival efficiency rate weighted average + departure efficiency rate weighted average

Example: ATL on 8/28/2005, for the quarter hour from 16:00 to 16:14:
Departures: 19
Departure Demand: 20
Departure Rate: 24
Departure Efficiency Rate: 19/20 = 95.00

Arrivals: 21
Arrival Demand: 42
Arrival Rate: 22
Arrival Efficiency Rate: 21/22 = 95.45

Total airport demand: departure demand + arrival demand = 62

SAER = (20/62) x 95.00 + (42/62) x 95.45 = 30.65 + 64.66 = 95.31

Figure 3: SAER Calculation

Date	CWSU Forecast	TAER	SAER	% On Time Arrivals	Weather Delays
5/14/10	18	97.45	97.46	64.20	70
5/15/10	70	96.70	95.81	74.07	48
5/16/10	65	95.12	93.74	72.10	85
6/1/10	50	93.75	90.35	59.57	138
6/2/10	80	96.62	96.52	80.08	27
6/3/10	265	96.77	95.14	82.52	13
6/4/10	75	96.56	95.67	78.44	14
6/5/10	65	96.52	97.38	90.79	0
6/6/10	70	91.34	89.60	64.39	88
6/7/10	-113	95.27	90.75	49.70	145
6/8/10	42	96.27	90.21	64.80	195
6/9/10	-120	90.77	90.26	38.05	193
6/18/10	-315	96.26	93.38	36.31	245
6/20/10	150	96.48	96.49	80.11	5
6/23/10	92	97.35	93.37	68.58	41
6/25/10	60	93.54	88.93	55.19	111
6/26/10	220	95.50	94.00	82.65	17
7/5/10	-117	96.56	95.32	75.96	11
7/6/10	59	98.27	92.75	66.73	85
7/7/10	35	98.54	95.45	70.60	46
7/8/10	5	97.74	91.11	61.64	63
7/9/10	55	97.70	91.57	70.35	41
7/10/10	-110	97.52	96.52	73.77	9
7/11/10	5	94.37	94.56	68.32	44
7/12/10	72	97.27	92.35	71.38	55
7/13/10	43	98.28	94.11	70.43	58
7/14/10	35	98.62	97.00	78.52	32
7/17/10	-25	98.54	97.64	83.13	4
7/18/10	45	98.23	98.41	82.26	0
7/20/10	40	98.46	96.98	63.50	87
7/21/10	100	97.65	97.93	79.96	10
7/22/10	13	98.78	94.38	70.61	60
7/23/10	-38	99.20	97.61	67.34	54
7/24/10	35	97.55	95.06	75.15	20
7/25/10	50	97.21	97.38	74.44	24
7/26/10	100	97.58	96.40	70.71	42
7/27/10	55	98.26	96.91	71.72	54
7/28/10	-30	97.47	91.39	67.83	71
7/29/10	-15	97.08	96.44	68.43	62
7/30/10	-10	95.62	95.37	65.59	60
7/31/10	53	98.48	93.25	70.04	32
8/1/10	-45	98.95	97.83	76.16	13
8/2/10	18	93.13	94.95	73.18	51
8/3/10	-12	98.41	96.63	74.45	45
8/4/10	-2	97.18	93.17	56.24	66
8/5/10	67	96.61	83.91	60.56	106
8/8/10	-20	98.37	92.93	71.51	44
8/9/10	45	97.74	93.79	66.91	85
8/10/10	10	99.13	98.05	73.15	64
8/11/10	-330	95.38	93.15	34.25	276
8/12/10	50	98.79	91.79	63.85	66
8/13/10	-25	96.48	96.15	64.44	69
8/16/10	30	96.25	93.37	59.38	71
8/17/10	-17	95.01	93.90	64.54	76
8/18/10	43	97.56	95.46	70.30	44
8/19/10	45	98.46	95.81	70.78	38
8/20/10	-73	95.88	93.96	60.93	69
8/21/10	-56	98.55	91.53	60.65	88
8/26/10	-75	94.41	95.95	75.14	36
8/27/10	-72	95.25	96.10	82.77	23
8/28/10	-5	94.58	90.02	57.67	85
8/31/10	-20	96.73	97.01	84.38	30
9/9/10	210	96.74	93.43	82.59	18
9/13/10	-7	98.60	96.73	60.19	83
9/14/10	43	96.93	96.07	70.63	43
9/15/10	-12	96.03	93.75	49.13	90
9/20/10	120	96.86	96.70	84.54	0
9/21/10	-92	94.83	91.64	54.40	95
10/1/10	0	95.81	94.88	61.58	88
10/2/10	20	95.84	93.13	83.33	31
10/3/10	-30	94.20	92.32	60.00	66
10/4/10	-200	95.17	93.23	60.32	90

Figure 4: CWSU ZOA Stratus Forecast Accuracy, Terminal and System Airport Efficiency Ratings, Percent on Time Arrivals, and Weather Delays for SFO

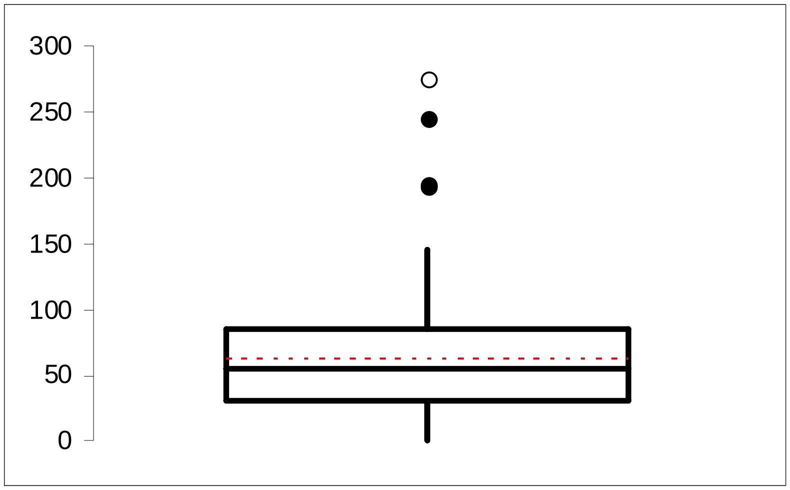

Figure 5: 2010 Data Sample - Weather Delays Interquartile Range Boxplot

Date	CWSU Forecast	TAER	SAER	% On Time Arrivals	Weather Delays
8/11/10	-330	95.38	93.15	34.25	276
6/18/10	-315	96.26	93.38	36.31	245
6/8/10	42	96.27	90.21	64.80	195
6/9/10	-120	90.77	90.26	38.05	193

Figure 6: Weather Delay Outliers – the Four Largest Daily Weather Delay Totals

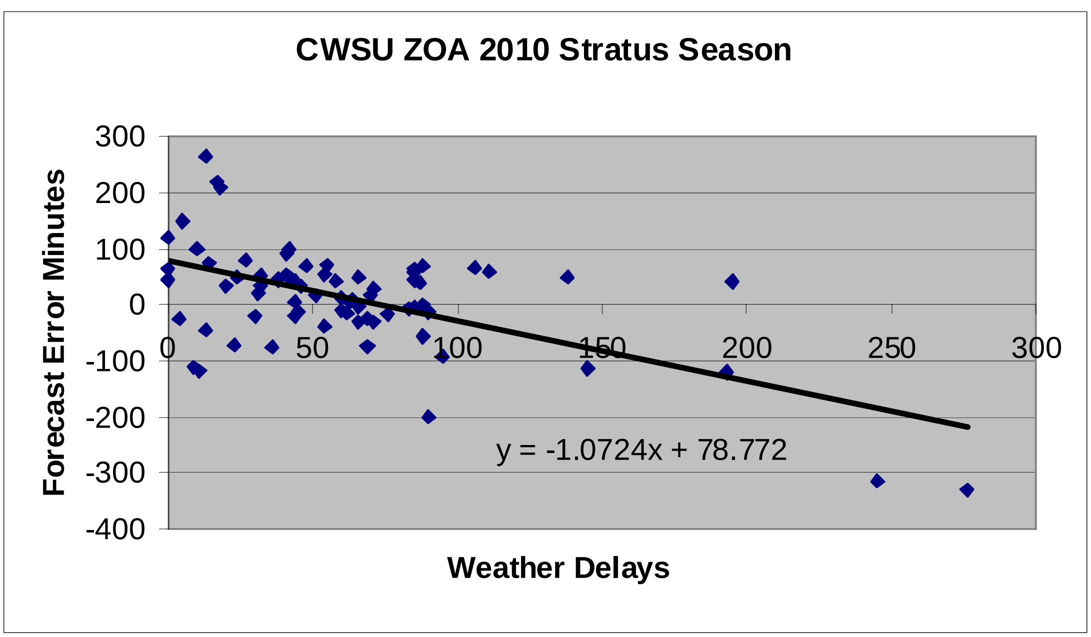

Figure 7: Weather Delays vs. CWSU Forecast Scatterplot with Trend Line

SUMMARY OUTPUT						
Regression Statistics						
Multiple R	0.58680269					
R Square	0.344337397					
Adjusted R Squa	0.334970788					
Standard Error	77.41258418					
Observations	72					
ANOVA						
	df	*SS*	*MS*	*F*	*Significance F*	
Regression	1	220305.3018	220305.3	36.76223	6.06E-08	
Residual	70	419489.5732	5992.708			
Total	71	639794.875				
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	78.77226833	14.3484661	5.489944	6.06E-07	50.15517846	107.3894
Weather Delays	-1.072449716	0.1768789	-6.06319	6.06E-08	-1.42522327	-0.719676

Figure 8: Regression statistics and analysis of variance for Weather Delays

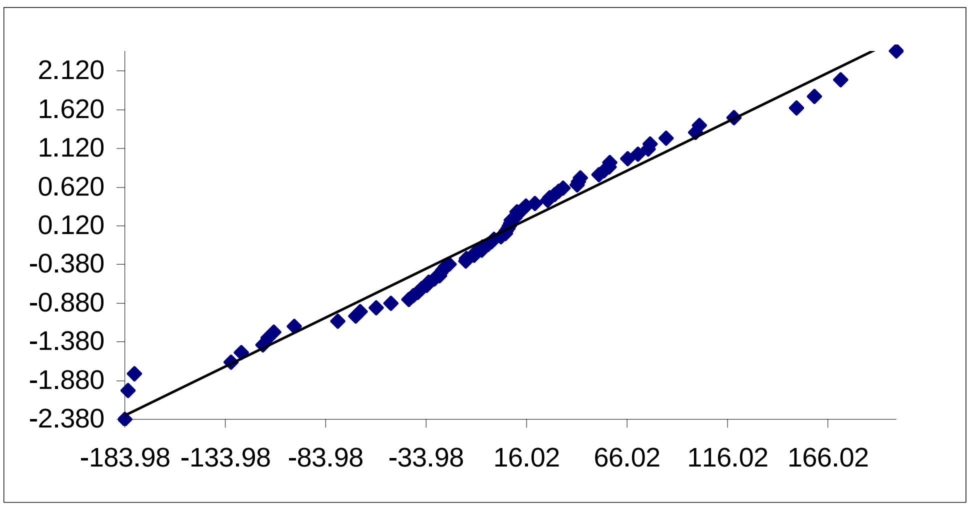

Figure 9: Plot of CWSU Forecast Regression Residuals

Pearson Correlations

	CWSU Forecast	On Time Arrivals	SAER	TAER	Weather Delays
CWSU Forecast	1.000	0.617	0.083	0.172	-0.587
On Time Arrivals		1.000	0.541	0.349	-0.875
SAER			1.000	0.451	-0.533
TAER				1.000	-0.373
Weather Delays					1.000

Pearson Probabilities

	CWSU Forecast	On Time Arrivals	SAER	TAER	Weather Delays
CWSU Forecast	-	0.000	0.488	0.149	0.000
On Time Arrivals		-	0.000	0.003	0.000
SAER			-	0.000	0.000
TAER				-	0.001
Weather Delays					-

Figure 10: Correlation Matrix (Significant Correlations in Red)

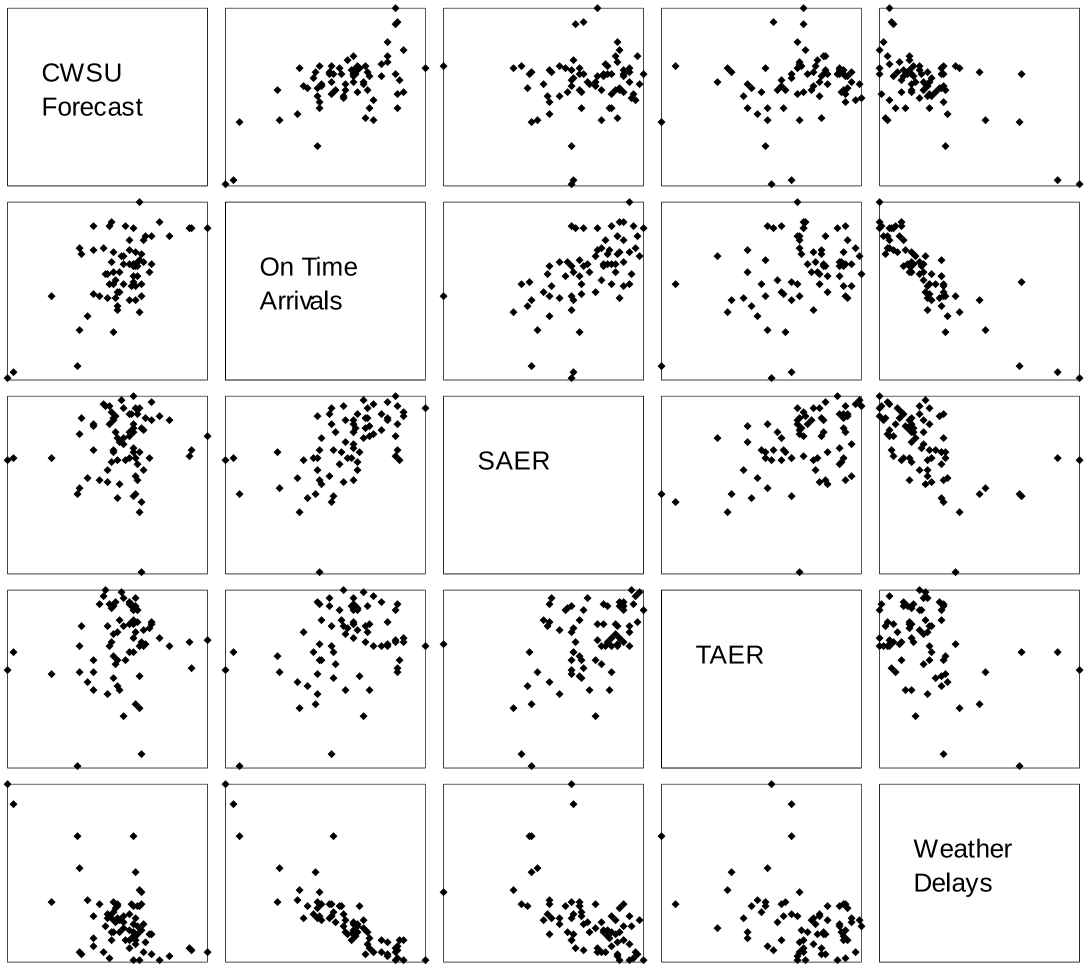

Figure 11: Scatterplot Matrix

SUMMARY OUTPUT						
Regression Statistics						
Multiple R	0.701077117					
R Square	0.491509124					
Adjusted R Square	0.46115146					
Standard Error	69.6826341					
Observations	72					
ANOVA						
	df	*SS*	*MS*	*F*	*Significance F*	
Regression	4	314465.0189	78616.25	16.19061	2.52737E-09	
Residual	67	325329.8561	4855.669			
Total	71	639794.875				
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	793.0795398	539.402916	1.470292	0.146164	-283.5734838	1869.7326
TAER	2.522592448	5.376775043	0.469165	0.640475	-8.209498417	13.254683
SAER	-14.00909076	3.907493838	-3.58519	0.000634	-21.80848325	-6.209698
% On Time Arrivals	4.809110484	1.562080576	3.078657	0.003013	1.691183744	7.9270372
Weather Delays	-0.525693648	0.334433992	-1.57189	0.120687	-1.193226874	0.1418396

Figure 12: Regression Stats and Analysis of Variance for All Coefficients

SUMMARY OUTPUT						
Regression Statistics						
Multiple R	0.685292707					
R Square	0.469626094					
Adjusted R Square	0.454252937					
Standard Error	70.12726597					
Observations	72					
ANOVA						
	df	*SS*	*MS*	*F*	*Significance F*	
Regression	2	300464.3682	150232.1841	30.54845	3.14836E-10	
Residual	69	339330.5068	4917.833432			
Total	71	639794.875				
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	721.8105529	319.1754735	2.261485023	0.026884	85.07310308	1358.548
SAER	-12.5370061	3.687580752	-3.399791608	0.001125	-19.89352613	-5.18049
% On Time Arrivals	6.90386418	0.889819296	7.758726084	5.52E-11	5.128723807	8.679005

Figure 13: Regression Statistics and Analysis of Variance for SAER and POTA Prediction of CWSU Forecast

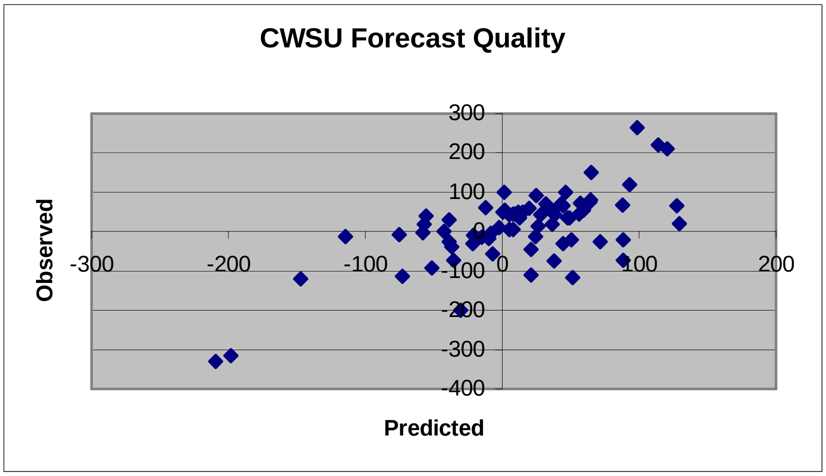

Figure 14: Scatterplot of Observed and Predicted CWSU Forecast Quality

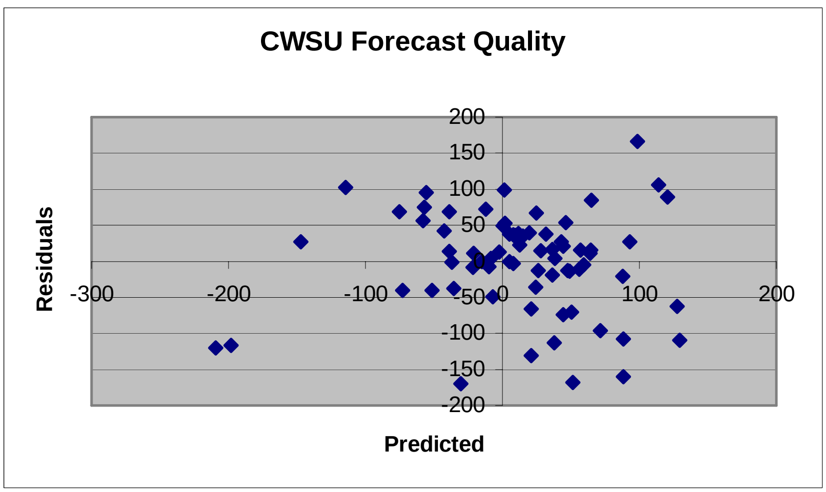

Figure 15: Scatterplot of Residuals and Predicted CWSU Forecast Scores

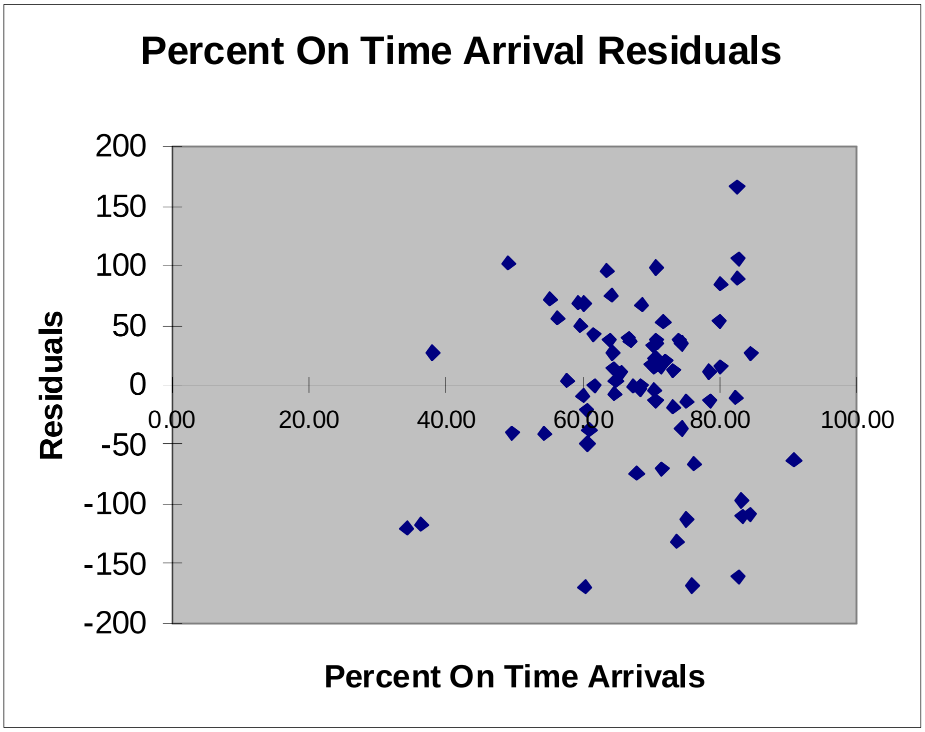

Figure 16: Percent on Time Arrival Residuals Plot

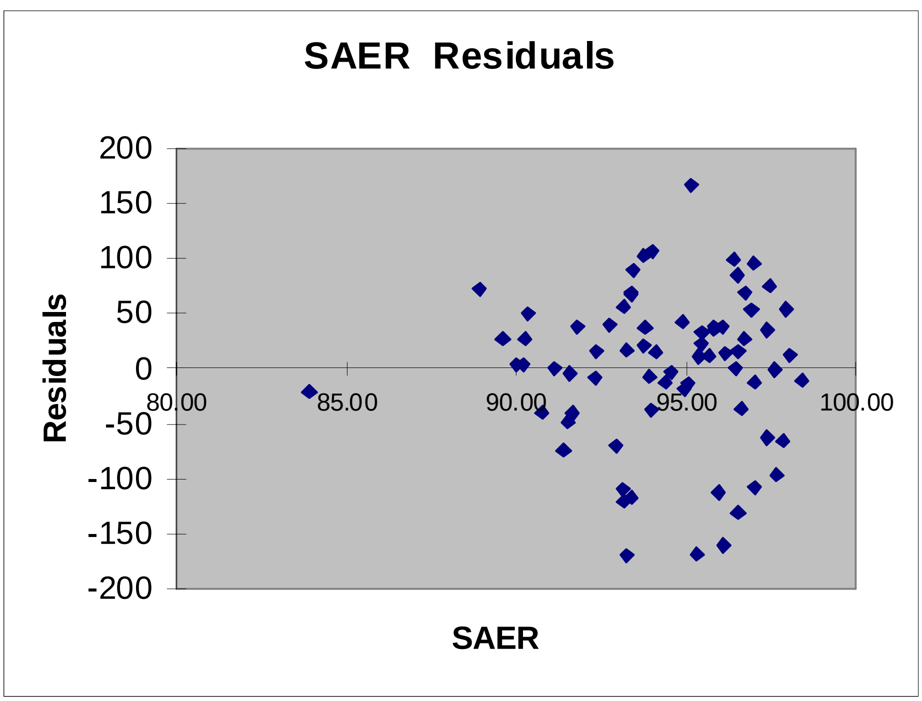

Figure 17: SAER Residual Plot

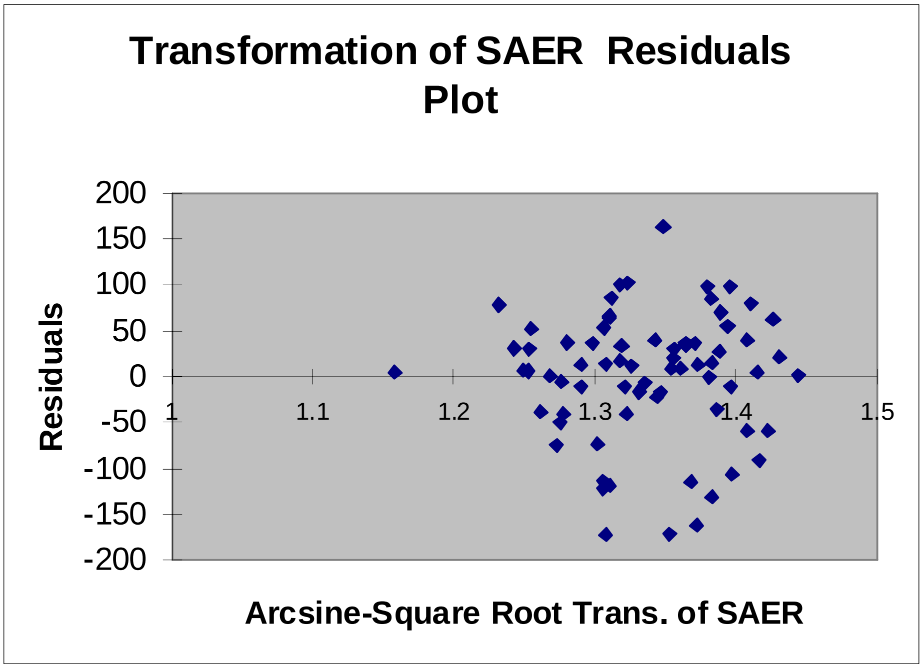

Figure 18: Transformed SAER Residual Plot

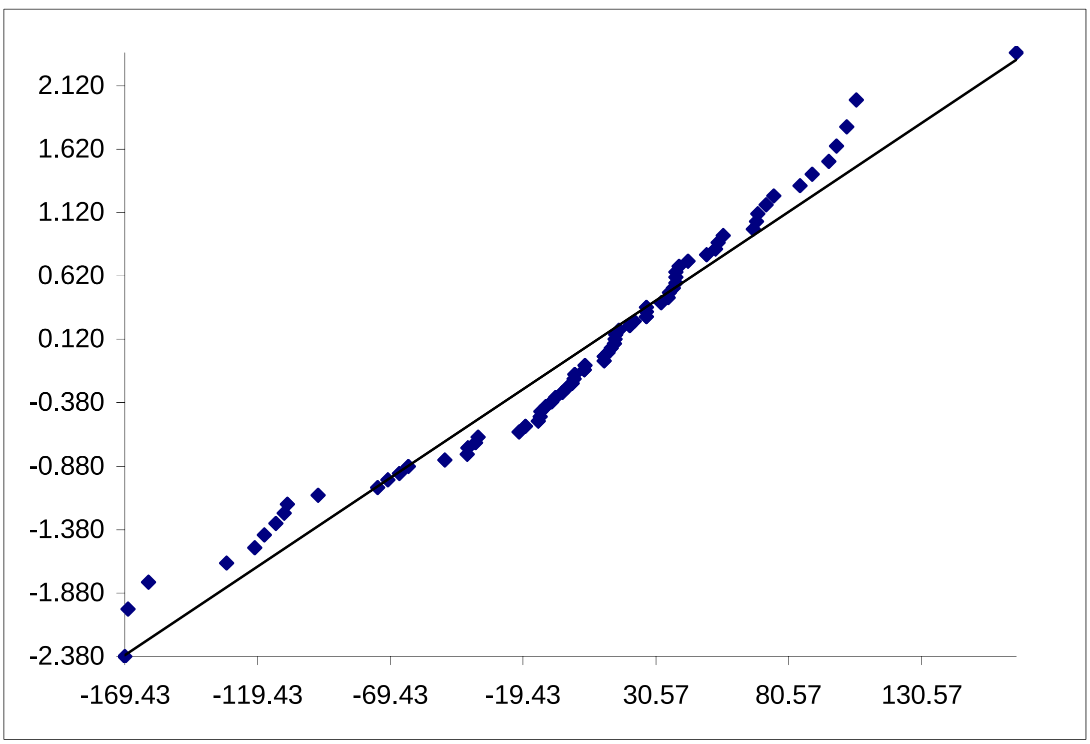

Figure 19: Normal P-plot of Residuals

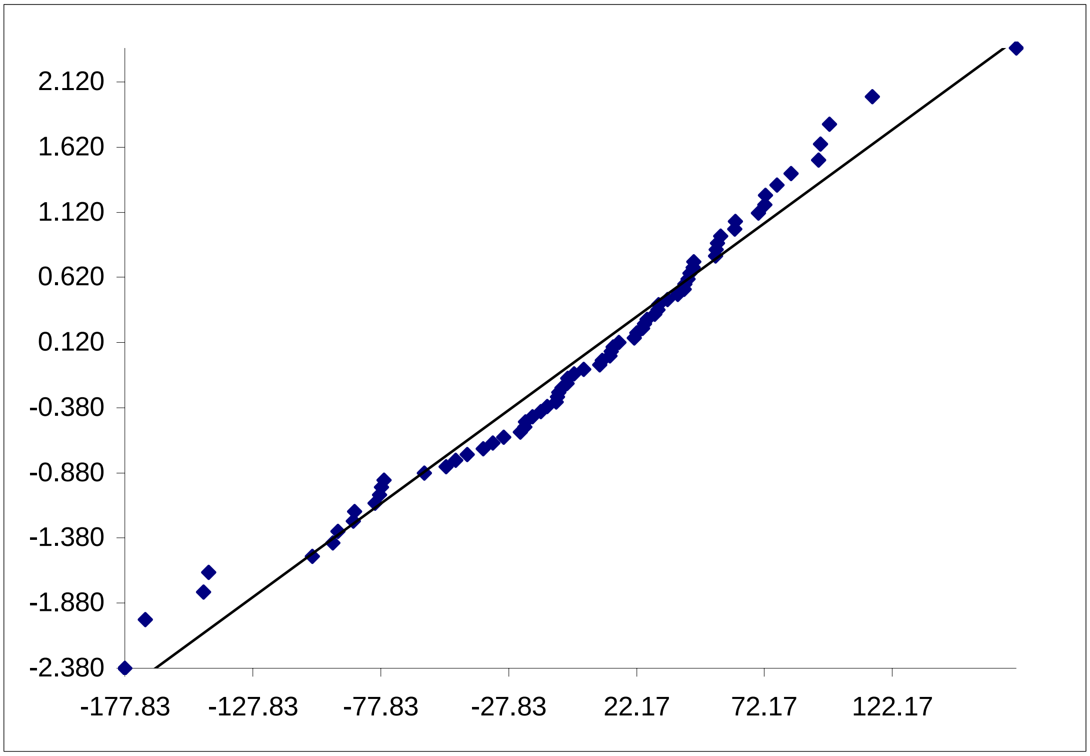

Figure 20: Normal P-plot of Full Regression Residuals